

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
3 January 2002 (03.01.2002)

PCT

(10) International Publication Number  
WO 02/00879 A2

(51) International Patent Classification<sup>7</sup>: C12N 15/12

(21) International Application Number: PCT/US01/20553

(22) International Filing Date: 27 June 2001 (27.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/214,358	28 June 2000 (28.06.2000)	US
60/215,638	30 June 2000 (30.06.2000)	US
60/279,997	30 March 2001 (30.03.2001)	US

(71) Applicant: GLYCOFI, INC. [US/US]; 34 South Main Street, Hanover, NH 03755 (US).

(72) Inventor: GERNGROSS, Tillman, U.; 16 Dorrance Place, Hanover, NH 03755 (US).

(74) Agents: PABST, Patrea, L. et al.; Holland & Knight LLP, One Atlantic Center, Suite 2000, 1201 West Peachtree Street, Atlanta, GA 30309-3400 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

(54) Title: METHODS FOR PRODUCING MODIFIED GLYCOPROTEINS

WO 02/00879

(57) **Abstract:** Cell lines having genetically modified glycosylation pathways that allow them to carry out a sequence of enzymatic reactions, which mimic the processing of glycoproteins in humans, have been developed. Recombinant proteins expressed in these engineered hosts yield glycoproteins more similar, if not substantially identical to their human counterparts. The lower eukaryotes, which ordinarily produce high-mannose containing *N*-glycans, including unicellular and multicellular fungi are modified to produce *N*-glycans such as Man<sub>5</sub>GlcNAc<sub>2</sub> or other structures along human glycosylation pathways. This is achieved using a combination of engineering and/or selection of strains which: do not express certain enzymes which create the undesirable complex structures characteristic of the fungal glycoproteins, which express exogenous enzymes selected either to have optimal activity under the conditions present in the fungi where activity is desired, or which are targeted to an organelle where optimal activity is achieved, and combinations thereof wherein the genetically engineered eukaryote expresses multiple exogenous enzymes required to produce "human-like" glycoproteins.

## METHODS FOR PRODUCING MODIFIED GLYCOPROTEINS

### Cross Reference To Related Applications

Priority is claimed to U.S. Provisional Application Serial No.

60/214,358, filed on June 28, 2000, U.S. Provisional Application Serial No.

5 60/215,638, filed on June 30, 2000, and U.S. Provisional Application Serial No. 60/279,997, filed on March 30, 2001.

### FIELD OF THE INVENTION

The present invention is directed to methods and compositions by which fungi or other eukaryotic microorganisms can be genetically modified

10 to produce glycosylated proteins (glycoproteins) having patterns of glycosylation similar to glycoproteins produced by animal cells, especially human cells, which are useful as human or animal therapeutic agents.

### BACKGROUND OF THE INVENTION

#### Glycosylation Pathways

15 *De novo* synthesized proteins may undergo further processing in cells, known as post-translational modification. In particular, sugar residues may be added enzymatically, a process known as glycosylation. The resulting proteins bearing covalently linked oligosaccharide side chains are known as glycosylated proteins or glycoproteins. Bacteria typically do not 20 glycosylate proteins; in cases where glycosylation does occur it usually occurs at nonspecific sites in the protein (Moens and Vanderleyden, Arch. Microbiol. 1997 168(3):169-175).

Eukaryotes commonly attach a specific oligosaccharide to the side chain of a protein asparagine residue, particularly an asparagine which 25 occurs in the sequence Asn-Xaa-Ser/Thr/Cys (where Xaa represents any amino acid). Following attachment of the saccharide moiety, known as an *N*-glycan, further modifications may occur *in vivo*. Typically these modifications occur via an ordered sequence of enzymatic reactions, known

as a cascade. Different organisms provide different glycosylation enzymes (glycosyltransferases and glycosidases) and different glycosyl substrates, so that the final composition of a sugar side chain may vary markedly depending upon the host.

5 For example, microorganisms such as filamentous fungi and yeast (lower eukaryotes) typically add additional mannose and/or mannosylphosphate sugars. The resulting glycan is known as a "high-mannose" type or a mannan. By contrast, in animal cells, the nascent oligosaccharide side chain may be trimmed to remove several mannose

10 residues and elongated with additional sugar residues that typically do not occur in the *N*-glycans of lower eukaryotes. See R.K. Brethauer, et al. *Biotechnology and Applied Biochemistry*, 1999, 30, 193-200; W. Martinet, et al. *Biotechnology Letters*, 1998, 20, 1171-1177; S. Weikert, et al. *Nature Biotechnology*, 1999, 17, 1116-1121; M. Malissard, et al. *Biochemical and Biophysical Research Communications*, 2000, 267, 169-173; Jarvis, et al. 1998 Engineering *N*-glycosylation pathways in the baculovirus-insect cell system, *Current Opinion in Biotechnology*, 9:528-533; and M. Takeuchi, 1997 *Trends in Glycoscience and Glycotechnology*, 1997, 9, S29-S35.

The *N*-glycans that are produced in humans and animals are 20 commonly referred to as complex *N*-glycans. A complex *N*-glycan means a structure with typically two to six outer branches with a sialyllactosamine sequence linked to an inner core structure  $\text{Man}_3\text{GlcNAc}_2$ . A complex *N*-glycan has at least one branch, and preferably at least two, of alternating GlcNAc and galactose (Gal) residues that terminate in oligosaccharides such 25 as, for example: NeuNAc-; NeuAc $\alpha$ 2-6GalNAc $\alpha$ 1-; NeuAc $\alpha$ 2-3Gal $\beta$ 1-3GalNAc $\alpha$ 1-; NeuAc $\alpha$ 2-3/6Gal $\beta$ 1-4GlcNAc $\beta$ 1-; GlcNAc $\alpha$ 1-4Gal $\beta$ 1- (mucins only); Fu $\alpha$ 1-2Gal $\beta$ 1-(blood group H). Sulfate esters can occur on galactose, GalNAc, and GlcNAc residues, and phosphate esters can occur on mannose residues. NeuAc (Neu: neuraminic acid; Ac:acetyl) can be O-acetylated or replaced by NeuGl (*N*-glycolylneuraminic acid). Complex *N*-glycans may also have intrachain substitutions of bisecting GlcNAc and core fucose (Fuc).

Human glycosylation begins with a sequential set of reactions in the endoplasmatic reticulum (ER) leading to a core oligosaccharide structure, which is transferred onto *de novo* synthesized proteins at the asparagine residue in the sequence Asn-Xaa-Ser/Thr (see Figure 1A). Further processing by glucosidases and mannosidases occurs in the ER before the nascent glycoprotein is transferred to the early Golgi apparatus, where additional mannose residues are removed by Golgi-specific 1,2-mannosidases. Processing continues as the protein proceeds through the Golgi. In the medial Golgi a number of modifying enzymes including *N*-acetylglucosamine transferases (GnT I, GnT II, GnT III, GnT IV GnT V GnT VI), mannosidase II, fucosyltransferases add and remove specific sugar residues (see Figure 1B). Finally in the trans Golgi, the *N*-glycans are acted on by galactosyl transferases and sialyltransferases (ST) and the finished glycoprotein is released from the Golgi apparatus. The protein *N*-glycans of animal glycoproteins have bi-, tri-, or tetra-antennary structures, and may typically include galactose, fucose, and *N*-acetylglucosamine. Commonly the terminal residues of the *N*-glycans consist of sialic acid. A typical structure of a human *N*-glycan is shown in Figure 1B.

#### Sugar Nucleotide Precursors

The *N*-glycans of animal glycoproteins typically include galactose, fucose, and terminal sialic acid. These sugars are not generally found on glycoproteins produced in yeast and filamentous fungi. In humans, the full range of nucleotide sugar precursors (e.g. UDP-*N*-acetylglucosamine, UDP-*N*-acetylgalactosamine, CMP-*N*-acetylneuraminic acid, UDP-galactose, GDP-fucose etc.) are generally synthesized in the cytosol and transported into the Golgi, where they are attached to the core oligosaccharide by glycosyltransferases. (Sommers and Hirschberg, 1981 *J. Cell Biol.* 91(2): A406-A406; Sommers and Hirschberg 1982 *J. Biol. Chem.* 257(18): 811-817; Perez and Hirschberg 1987 *Methods in Enzymology* 138: 709-715).

Glycosyl transfer reactions typically yield a side product which is a nucleoside diphosphate or monophosphate. While monophosphates can be directly exported in exchange for nucleoside triphosphate sugars by an

antiport mechanism, diphosphonucleosides (e.g. GDP) have to be cleaved by phosphatases (e.g. GDFPase) to yield nucleoside monophosphates and inorganic phosphate prior to being exported. This reaction is important for efficient glycosylation; for example, GDFPase from *S. cerevisiae* has been found to be necessary for mannosylation. However the GDFPase has 90% reduced activity toward UDP (Berninsone et al., 1994 *J. Biol. Chem.* 269(1):207-211 $\alpha$ ). Lower eukaryotes typically lack UDP-specific diphosphatase activity in the Golgi since they do not utilize UDP-sugar precursors for Golgi-based glycoprotein synthesis. *Schizosaccharomyces pombe*, a yeast found to add galactose residues to cell wall polysaccharides (from UDP-galactose) has been found to have specific UDPase activity, indicating the requirement for such an enzyme (Berninsone et al., 1994). UDP is known to be a potent inhibitor of glycosyltransferases and the removal of this glycosylation side product is important in order to prevent glycosyltransferase inhibition in the lumen of the Golgi (Khatara et al., 1974). See Berninsone, P., et al. 1995. *J. Biol. Chem.* 270(24): 14564-14567; Beaudet, L., et al. 1998 *Abc Transporters: Biochemical, Cellular, and Molecular Aspects.* 292: 397-413.

#### Compartmentalization of Glycosylation Enzymes

Glycosyltransferases and mannosidases line the inner (luminal) surface of the ER and Golgi apparatus and thereby provide a catalytic surface that allows for the sequential processing of glycoproteins as they proceed through the ER and Golgi network. The multiple compartments of the cis, medial, and trans Golgi and the trans Golgi Network (TGN), provide the different localities in which the ordered sequence of glycosylation reactions can take place. As a glycoprotein proceeds from synthesis in the ER to full maturation in the late Golgi or TGN, it is sequentially exposed to different glycosidases, mannosidases and glycosyltransferases such that a specific N-glycan structure may be synthesized. The enzymes typically include a catalytic domain, a stem region, a membrane spanning region and an N-terminal cytoplasmic tail. The latter three structural components are responsible for directing a glycosylation enzyme to the appropriate locus.

Localization sequences from one organism may function in other organisms. For example the membrane spanning region of  $\alpha$ -2,6-sialyltransferase ( $\alpha$ -2,6-ST) from rats, an enzyme known to localize in the rat trans Golgi, was shown to also localize a reporter gene (invertase) in the yeast Golgi (Schwientek, et al., 1995). However, the very same membrane spanning region as part of a full-length of  $\alpha$ -2,6-sialyltransferase was retained in the ER and not further transported to the Golgi of yeast (Krezdorn et al., 1994). A full length GalT from humans was not even synthesized in yeast, despite demonstrably high transcription levels. On the other hand the transmembrane region of the same human GalT fused to an invertase reporter was able to direct localization to the yeast Golgi, albeit at low production levels. Schwientek and co-workers have shown that fusing 28 amino acids of a yeast mannosyltransferase (Mnt1), a region containing an N-terminal cytoplasmic tail, a transmembrane region and eight amino acids of the stem region, to the catalytic domain of human GalT are sufficient for Golgi localization of an active GalT (Schwientek et al. 1995 J. Biol. Chem. 270(10):5483-5489). Other galactosyltransferases appear to rely on interactions with enzymes resident in particular organelles since after removal of their transmembrane region they are still able to localize properly.

Improper localization of a glycosylation enzyme may prevent proper functioning of the enzyme in the pathway. For example *Aspergillus nidulans*, which has numerous  $\alpha$ -1,2-mannosidases (Eades and Hintz, 2000 Gene 255(1):25-34), does not add GlcNAc to  $\text{Man}_5\text{GlcNAc}_2$  when transformed with the rabbit GnT I gene, despite a high overall level of GnT I activity (Kalsner et al., 1995). GnT I, although actively expressed, may be incorrectly localized such that the enzyme is not in contact with both of its substrates: the nascent *N*-glycan of the glycoprotein and UDP-GlcNAc. Alternatively, the host organism may not provide an adequate level of UDP-GlcNAc in the Golgi.

Glycoproteins Used Therapeutically

A significant fraction of proteins isolated from humans or other animals are glycosylated. Among proteins used therapeutically, about 70% are glycosylated. If a therapeutic protein is produced in a microorganism 5 host such as yeast, however, and is glycosylated utilizing the endogenous pathway, its therapeutic efficiency is typically greatly reduced. Such glycoproteins are typically immunogenic in humans and show a reduced half-life *in vivo* after administration (Takeuchi, 1997).

Specific receptors in humans and animals can recognize terminal 10 mannose residues and promote the rapid clearance of the protein from the bloodstream. Additional adverse effects may include changes in protein folding, solubility, susceptibility to proteases, trafficking, transport, compartmentalization, secretion, recognition by other proteins or factors, 15 antigenicity, or allergenicity. Accordingly, it has been necessary to produce therapeutic glycoproteins in animal host systems, so that the pattern of glycosylation is identical or at least similar to that in humans or in the intended recipient species. In most cases a mammalian host system, such as mammalian cell culture, is used.

Systems for Producing Therapeutic Glycoproteins

20 In order to produce therapeutic proteins that have appropriate glycoforms and have satisfactory therapeutic effects, animal or plant-based expression systems have been used. The available systems include:

1. Chinese hamster ovary cells (CHO), mouse fibroblast cells and mouse myeloma cells (Arzneimittelforschung. 1998 Aug;48(8):870-25 880);
2. transgenic animals such as goats, sheep, mice and others (Dente Prog. Clin. Biol. 1989 Res. 300:85-98, Ruther et al., 1988 Cell 53(6):847-856; Ware, J., et al. 1993 *Thrombosis and Haemostasis* 69(6): 1194-1194; Cole, E. S., et al. 1994 *J.Cell.Biochem.* 265-265);
- 30 3. plants (*Arabidopsis thaliana*, tobacco etc.) (Staub, et al. 2000 *Nature Biotechnology* 18(3): 333-338) (McGarvey, P. B., et al. 1995 *Bio-*

*Technology* 13(13): 1484-1487; Bardor, M., et al. 1999 *Trends in Plant Science* 4(9): 376-380);

4. insect cells (*Spodoptera frugiperda* Sf9, Sf21, *Trichoplusia ni*, etc. in combination with recombinant baculoviruses such as *Autographa californica* multiple nuclear polyhedrosis virus which infects lepidopteran cells) (Altmans et al., 1999 *Glycoconj. J.* 16(2):109-123).

Recombinant human proteins expressed in the above-mentioned host systems may still include non-human glycoforms (Raju et al., 2000 *Annals*

10 *Biochem.* 283(2):123-132). In particular, fraction of the *N*-glycans may lack terminal sialic acid, typically found in human glycoproteins. Substantial efforts have been directed to developing processes to obtain glycoproteins that are as close as possible in structure to the human forms, or have other therapeutic advantages. Glycoproteins having specific glycoforms may be 15 especially useful, for example in the targeting of therapeutic proteins. For example, the addition of one or more sialic acid residues to a glycan side chain may increase the lifetime of a therapeutic glycoprotein *in vivo* after administration. Accordingly, the mammalian host cells may be genetically engineered to increase the extent of terminal sialic acid in glycoproteins 20 expressed in the cells. Alternatively sialic acid may be conjugated to the protein of interest *in vitro* prior to administration using a sialic acid transferase and an appropriate substrate. In addition, changes in growth medium composition or the expression of enzymes involved in human glycosylation have been employed to produce glycoproteins more closely 25 resembling the human forms (S. Weikert, et al., *Nature Biotechnology*, 1999, 17, 1116-1121; Werner, Noe, et al 1998 *Arzneimittelforschung* 48(8):870-880; Weikert, Papac et al., 1999; Andersen and Goochee 1994 *Cur. Opin.Biotechnol.* 5: 546-549; Yang and Butler 2000 *Biotechnol.Bioengin.* 68(4): 370-380). Alternatively cultured human cells 30 may be used.

However, all of the existing systems have significant drawbacks. Only certain therapeutic proteins are suitable for expression in animal or

plant systems (e.g. those lacking in any cytotoxic effect or other effect adverse to growth). Animal and plant cell culture systems are usually very slow, frequently requiring over a week of growth under carefully controlled conditions to produce any useful quantity of the protein of interest. Protein 5 yields nonetheless compare unfavorably with those from microbial fermentation processes. In addition cell culture systems typically require complex and expensive nutrients and cofactors, such as bovine fetal serum. Furthermore growth may be limited by programmed cell death (apoptosis).

Moreover, animal cells (particularly mammalian cells) are highly 10 susceptible to viral infection or contamination. In some cases the virus or other infectious agent may compromise the growth of the culture, while in other cases the agent may be a human pathogen rendering the therapeutic protein product unfit for its intended use. Furthermore many cell culture processes require the use of complex, temperature-sensitive, animal-derived 15 growth media components, which may carry pathogens such as bovine spongiform encephalopathy (BSE) prions. Such pathogens are difficult to detect and/or difficult to remove or sterilize without compromising the growth medium. In any case, use of animal cells to produce therapeutic proteins necessitates costly quality controls to assure product safety.

20 Transgenic animals may also be used for manufacturing high-volume therapeutic proteins such as human serum albumin, tissue plasminogen activator, monoclonal antibodies, hemoglobin, collagen, fibrinogen and others. While transgenic goats and other transgenic animals (mice, sheep, cows, etc.) can be genetically engineered to produce therapeutic proteins at 25 high concentrations in the milk, the process is costly since every batch has to undergo rigorous quality control. Animals may host a variety of animal or human pathogens, including bacteria, viruses, fungi, and prions. In the case of scrapies and bovine spongiform encephalopathy, testing can take about a year to rule out infection. The production of therapeutic compounds is thus 30 preferably carried out in a well-controlled sterile environment, e.g. under Good Manufacturing Practice (GMP) conditions. However, it is not generally feasible to maintain animals in such environments. Moreover,

whereas cells grown in a fermenter are derived from one well characterized Master Cell Bank (MCB), transgenic animal technology relies on different animals and thus is inherently non-uniform. Furthermore external factors such as different food uptake, disease and lack of homogeneity within a herd, 5 may effect glycosylation patterns of the final product. It is known in humans, for example, that different dietary habits result in differing glycosylation patterns.

Transgenic plants have been developed as a potential source to obtain proteins of therapeutic value. However, high level expression of proteins in 10 plants suffers from gene silencing, a mechanism by which the genes for highly expressed proteins are down-regulated in subsequent plant generations. In addition, plants add xylose and/or  $\alpha$ -1,3-linked fucose to protein *N*-glycans, resulting in glycoproteins that differ in structure from animals and are immunogenic in mammals (Altmann, Marz et al., 1995 15 Glycoconj. J. 12(2):150-155). Furthermore, it is generally not practical to grow plants in a sterile or GMP environment, and the recovery of proteins from plant tissues is more costly than the recovery from fermented microorganisms.

#### Glycoprotein Production Using Eukaryotic Microorganisms

20 The lack of a suitable expression system is thus a significant obstacle to the low-cost and safe production of recombinant human glycoproteins. Production of glycoproteins via the fermentation of microorganisms would offer numerous advantages over the existing systems. For example, 25 fermentation-based processes may offer (a) rapid production of high concentrations of protein; (b) the ability to use sterile, well-controlled production conditions (e.g. GMP conditions); (c) the ability to use simple, chemically defined growth media; (d) ease of genetic manipulation; (e) the absence of contaminating human or animal pathogens; (f) the ability to express a wide variety of proteins, including those poorly expressed in cell 30 culture owing to toxicity etc.; (g) ease of protein recovery (e.g. via secretion into the medium). In addition, fermentation facilities are generally far less costly to construct than cell culture facilities.

As noted above, however, bacteria, including species such as *Escherichia coli* commonly used to produce recombinant proteins, do not glycosylate proteins in a specific manner like eukaryotes. Various methylotrophic yeasts such as *Pichia pastoris*, *Pichia methanolica*, and 5 *Hansenula polymorpha*, are particularly useful as eukaryotic expression systems, since they are able to grow to high cell densities and/or secrete large quantities of recombinant protein. However, as noted above, glycoproteins expressed in these eukaryotic microorganisms differ substantially in 10 *N*-glycan structure from those in animals. This has prevented the use of yeast or filamentous fungi as hosts for the production of many useful glycoproteins.

Several efforts have been made to modify the glycosylation pathways of eukaryotic microorganisms to provide glycoproteins more suitable for use as mammalian therapeutic agents. For example, several glycosyltransferases 15 have been separately cloned and expressed in *S. cerevisiae* (GalT, GnT I), *Aspergillus nidulans* (GnT I) and other fungi (Yoshida et al., 1999, Kalsner et al., 1995 *Glycoconj. J.* 12(3):360-370, Schwientek et al., 1995). However, *N*-glycans with human characteristics were not obtained.

Yeast produce a variety of mannosyltransferases e.g. 1,3- 20 mannosyltransferases (e.g. MNN1 in *S. cerevisiae*) (Graham and Emr, 1991 *J. Cell. Biol.* 114(2):207-218), 1,2-mannosyltransferases (e.g. KTR/KRE family from *S. cerevisiae*), 1,6-mannosyltransferases (OCH1 from *S. cerevisiae*), mannosylphosphate transferases (MNN4 and MNN6 from *S. cerevisiae*) and additional enzymes that are involved in endogenous 25 glycosylation reactions. Many of these genes have been deleted individually, giving rise to viable organisms having altered glycosylation profiles. Examples are shown in Table 1.

Table 1. Examples of yeast strains having altered mannosylation

Strain	<i>N</i> -glycan (wild type)	Mutation	<i>N</i> -glycan (mutant)	Reference
<i>S. pombe</i>	Man <sub>9</sub> GlcNAc <sub>2</sub>	OCH1	Man <sub>8</sub> GlcNAc <sub>2</sub>	Yoko-o et al., 2001 FEBS Lett. 489(1):75-80
<i>S. cerevisiae</i>	Man <sub>9</sub> GlcNAc <sub>2</sub>	OCH1/MNN1	Man <sub>8</sub> GlcNAc <sub>2</sub>	Nakanishi-Shindo et al., 1993 J. Biol. Chem. 268(35):26338- 26345
<i>S. cerevisiae</i>	Man <sub>9</sub> GlcNAc <sub>2</sub>	OCH1/MNN1/ MNN4	Man <sub>8</sub> GlcNAc <sub>2</sub>	Chiba et al., 1998 J. Biol. Chem. 273, 26298-26304

In addition, Japanese Patent Application Public No. 8-336387 discloses an *OCH1* mutant strain of *Pichia pastoris*. The *OCH1* gene encodes 1,6-mannosyltransferase, which adds a mannose to the glycan structure Man<sub>8</sub>GlcNAc<sub>2</sub> to yield Man<sub>9</sub>GlcNAc<sub>2</sub>. The Man<sub>9</sub>GlcNAc<sub>2</sub> structure is then a substrate for further mannosylation *in vivo*, leading to the hypermannosylated glycoproteins that are characteristic of yeasts and typically may have at least 30-40 mannose residue per *N*-glycan. In the *OCH1* mutant strain, proteins glycosylated with Man<sub>8</sub>GlcNAc<sub>2</sub> are accumulated and hypermannosylation does not occur. However, the structure Man<sub>8</sub>GlcNAc<sub>2</sub> is not a substrate for animal glycosylation enzymes, such as human UDP-GlcNAc transferase I, and accordingly the method is not useful for producing proteins with human glycosylation patterns.

Martinet et al. (Biotechnol. Lett. 1998, 20(12), 1171-1177) reported the expression of  $\alpha$ -1,2-mannosidase from *Trichoderma reesei* in *P. pastoris*. Some mannose trimming from the *N*-glycans of a model protein was observed. However, the model protein had no *N*-glycans with the structure Man<sub>5</sub>GlcNAc<sub>2</sub>, which would be necessary as an intermediate for the generation of complex *N*-glycans. Accordingly the method is not useful for producing proteins with human or animal glycosylation patterns.

Similarly, Chiba et al. 1998 expressed  $\alpha$ -1,2-mannosidase from *Aspergillus saitoi* in the yeast *Saccharomyces cerevisiae*. A signal peptide sequence (His-Asp-Glu-Leu) was engineered into the exogenous mannosidase to promote retention in the endoplasmic reticulum. In addition, 5 the yeast host was a mutant lacking three enzyme activities associated with hypermannosylation of proteins: 1,6-mannosyltransferase (*OCH1*); 1,3-mannosyltransferase (*MNN1*); and mannosylphosphatetrasferase (*MNN4*). The *N*-glycans of the triple mutant host thus consisted of the structure  $\text{Man}_8\text{GlcNAc}_2$ , rather than the high mannose forms found in wild-type *S.* 10 *cerevisiae*. In the presence of the engineered mannosidase, the *N*-glycans of a model protein (carboxypeptidase Y) were trimmed to give a mixture consisting of 27 mole %  $\text{Man}_5\text{GlcNAc}_2$ , 22 mole %  $\text{Man}_6\text{GlcNAc}_2$ , 22 mole %  $\text{Man}_7\text{GlcNAc}_2$ , 29 mole %  $\text{Man}_8\text{GlcNAc}_2$ . Trimming of the endogenous cell wall glycoproteins was less efficient, only 10 mole % of the *N*-glycans 15 having the desired  $\text{Man}_5\text{GlcNAc}_2$  structure.

Since only the  $\text{Man}_5\text{GlcNAc}_2$  glycans would be susceptible to further enzymatic conversion to human glycoforms, the method is not efficient for the production of proteins having human glycosylation patterns. In proteins having a single *N*-glycosylation site, at least 73 mole % would have an 20 incorrect structure. In proteins having two or three *N*-glycosylation sites, respectively at least 93 or 98 mole % would have an incorrect structure. Such low efficiencies of conversion are unsatisfactory for the production of therapeutic agents, particularly as the separation of proteins having different glycoforms is typically costly and difficult.

25 With the object of providing a more human-like glycoprotein derived from a fungal host, U.S. Patent No. 5,834,251 to Maras and Contreras discloses a method for producing a hybrid glycoprotein derived from *Trichoderma reesei*. A hybrid *N*-glycan has only mannose residues on the  $\text{Man}\alpha 1-6$  arm of the core and one or two complex antennae on the  $\text{Man}\alpha 1-3$  30 arm. While this structure has utility, the method has the disadvantage that numerous enzymatic steps must be performed *in vitro*, which is costly and time-consuming. Isolated enzymes are expensive to prepare and maintain,

may need unusual and costly substrates (e.g. UDP-GlcNAc), and are prone to loss of activity and/or proteolysis under the conditions of use.

It is therefore an object of the present invention to provide a system and methods for humanizing glycosylation of recombinant glycoproteins

5 expressed in *Pichia pastoris* and other lower eukaryotes such as *Hansenula polymorpha*, *Pichia stipitis*, *Pichia methanolica*, *Pichia* sp., *Kluyveromyces* sp., *Candida albicans*, *Aspergillus nidulans*, and *Trichoderma reseei*.

## SUMMARY OF THE INVENTION

Cell lines having genetically modified glycosylation pathways that

10 allow them to carry out a sequence of enzymatic reactions, which mimic the processing of glycoproteins in humans, have been developed. Recombinant proteins expressed in these engineered hosts yield glycoproteins more similar, if not substantially identical, to their human counterparts. The lower eukaryotes, which ordinarily produce high-mannose containing *N*-glycans,

15 including unicellular and multicellular fungi such as *Pichia pastoris*, *Hansenula polymorpha*, *Pichia stipitis*, *Pichia methanolica*, *Pichia* sp., *Kluyveromyces* sp., *Candida albicans*, *Aspergillus nidulans*, and *Trichoderma reseei*, are modified to produce *N*-glycans such as  $\text{Man}_5\text{GlcNAc}_2$  or other structures along human glycosylation pathways. This is achieved using a

20 combination of engineering and/or selection of strains which: do not express certain enzymes which create the undesirable complex structures characteristic of the fungal glycoproteins, which express exogenous enzymes selected either to have optimal activity under the conditions present in the fungi where activity is desired, or which are targeted to an organelle where

25 optimal activity is achieved, and combinations thereof wherein the genetically engineered eukaryote expresses multiple exogenous enzymes required to produce "human-like" glycoproteins.

In a first embodiment, the microorganism is engineered to express an exogenous  $\alpha$ -1,2-mannosidase enzyme having an optimal pH between 5.1

30 and 8.0, preferably between 5.9 and 7.5. In an alternative preferred embodiment, the exogenous enzyme is targeted to the endoplasmic reticulum

or Golgi apparatus of the host organism, where it trims *N*-glycans such as  $\text{Man}_8\text{GlcNAc}_2$  to yield  $\text{Man}_5\text{GlcNAc}_2$ . The latter structure is useful because it is identical to a structure formed in mammals, especially humans; it is a substrate for further glycosylation reactions *in vivo* and/or *in vitro* that

5 produce a finished *N*-glycan that is similar or identical to that formed in mammals, especially humans; and it is not a substrate for hypermannosylation reactions that occur *in vivo* in yeast and other microorganisms and that render a glycoprotein highly immunogenic in animals.

10 In a second embodiment, the glycosylation pathway of an eukaryotic microorganism is modified by (a) constructing a DNA library including at least two genes encoding exogenous glycosylation enzymes; (b) transforming the microorganism with the library to produce a genetically mixed population expressing at least two distinct exogenous glycosylation

15 enzymes; (c) selecting from the population a microorganism having the desired glycosylation phenotype. In a preferred embodiment, the DNA library includes chimeric genes each encoding a protein localization sequence and a catalytic activity related to glycosylation. Organisms modified using the method are useful for producing glycoproteins having a

20 glycosylation pattern similar or identical to mammals, especially humans.

In a third embodiment, the glycosylation pathway is modified to express a sugar nucleotide transporter enzyme. In a preferred embodiment, a nucleotide diphosphatase enzyme is also expressed. The transporter and diphosphatase improve the efficiency of engineered glycosylation steps, by

25 providing the appropriate substrates for the glycosylation enzymes in the appropriate compartments, reducing competitive product inhibition, and promoting the removal of nucleoside diphosphates.

#### DESCRIPTION OF THE FIGURES

Figure 1A is a schematic diagram of typical fungal *N*-glycosylation

30 pathway.

Figure 1B is a schematic diagram of a typical human *N*-glycosylation pathway.

#### DETAILED DESCRIPTION OF THE INVENTION

The methods and recombinant lower eukaryotic strains described

5 herein are used to make "humanized glycoproteins". The recombinant lower eukaryotes are made by engineering lower eukaryotes which do not express one or more enzymes involved in production of high mannose structures to express the enzymes required to produce human-like sugars. As used herein, a lower eukaryote is a unicellular or filamentous fungus. As used herein, a

10 "humanized glycoprotein" refers to a protein having attached thereto N-glycans including less than four mannose residues, and the synthetic intermediates (which are also useful and can be manipulated further *in vitro*) having at least five mannose residues. In a preferred embodiment, the glycoproteins produced in the recombinant lower eukaryotic strains contain

15 at least 27 mole % of the Man5 intermediate. This is achieved by cloning in a better mannosidase, i.e., an enzyme selected to have optimal activity under the conditions present in the organisms at the site where proteins are glycosylated, or by targeting the enzyme to the organelle where activity is desired.

20 In a preferred embodiment, eukaryotic strains which do not express one or more enzymes involved in the production of high mannose structures are used. These strains can be engineered or one of the many such mutants already described in yeasts, including a hypermannosylation-minus (OCH1) mutant in *Pichia pastoris*.

25 The strains can be engineered one enzyme at a time, or a library of genes encoding potentially useful enzymes can be created, and those strains having enzymes with optimal activities or producing the most "human-like" glycoproteins, selected.

Lower eukaryotes that are able to produce glycoproteins having the

30 attached *N*-glycan  $\text{Man}_5\text{GlcNAc}_2$  are particularly useful since (a) lacking a high degree of mannosylation (e.g. greater than 8 mannoses per *N*-glycan, or

especially 30-40 mannoses), they show reduced immunogenicity in humans; and (b) the *N*-glycan is a substrate for further glycosylation reactions to form an even more human-like glycoform, e.g. by the action of GlcNAc transferase I to form GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. Man<sub>5</sub>GlcNAc<sub>2</sub> must be formed 5 *in vivo* in a high yield, at least transiently, since all subsequent glycosylation reactions require Man<sub>5</sub>GlcNAc<sub>2</sub> or a derivative thereof. Accordingly, a yield is obtained of greater than 27 mole %, more preferably a yield of 50-100 mole %, glycoproteins in which a high proportion of *N*-glycans have Man<sub>5</sub>GlcNAc<sub>2</sub>. It is then possible to perform further glycosylation reactions 10 *in vitro*, using for example the method of U.S. Patent No. 5,834,251 to Maras and Contreras. In a preferred embodiment, at least one further glycosylation reaction is performed *in vivo*. In a highly preferred embodiment thereof, active forms of glycosylating enzymes are expressed in the endoplasmic reticulum and/or Golgi apparatus.

15 **Host Microorganisms**

Yeast and filamentous fungi have both been successfully used for the production of recombinant proteins, both intracellular and secreted (Cereghino, J. L. and J. M. Cregg 2000 *FEMS Microbiology Reviews* 24(1): 45-66; Harkki, A., et al. 1989 *Bio-Technology* 7(6): 596; Berka, R. M., et al. 20 1992 *Abstr. Papers Amer. Chem. Soc.* 203: 121-BIOT; Svetina, M., et al. 2000 *J. Biotechnol.* 76(2-3): 245-251.

Although glycosylation in yeast and fungi is very different than in humans, some common elements are shared. The first step, the transfer of the core oligosaccharide structure to the nascent protein, is highly conserved 25 in all eukaryotes including yeast, fungi, plants and humans (compare Figures 1A and 1B). Subsequent processing of the core oligosaccharide, however, differs significantly in yeast and involves the addition of several mannose sugars. This step is catalyzed by mannosyltransferases residing in the Golgi (e.g. OCH1, MNT1, MNN1, etc.), which sequentially add mannose sugars to 30 the core oligosaccharide. The resulting structure is undesirable for the production of humanoid proteins and it is thus desirable to reduce or eliminate mannosyl transferase activity. Mutants of *S. cerevisiae*, deficient

in mannosyl transferase activity (e.g. *och1* or *mnn9* mutants) have shown to be non-lethal and display a reduced mannose content in the oligosaccharide of yeast glycoproteins. Other oligosaccharide processing enzymes, such as mannosylphosphate transferase may also have to be eliminated depending on 5 the host's particular endogenous glycosylation pattern. After reducing undesired endogenous glycosylation reactions the formation of complex N-glycans has to be engineered into the host system. This requires the stable expression of several enzymes and sugar-nucleotide transporters. Moreover, one has to locate these enzymes in a fashion such that a sequential processing 10 of the maturing glycosylation structure is ensured.

#### Target Glycoproteins

The methods described herein are useful for producing glycoproteins, especially glycoproteins used therapeutically in humans. Such therapeutic proteins are typically administered by injection, orally, pulmonary, or other 15 means.

Examples of suitable target glycoproteins include, without limitation: erythropoietin, cytokines such as interferon- $\alpha$ , interferon- $\beta$ , interferon- $\gamma$ , interferon- $\omega$ , and granulocyte-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, soluble IgE receptor  $\alpha$ -chain, IgG, IgM, 20 urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, and osteoprotegerin.

#### Method for Producing Glycoproteins Comprising the *N*-glycan

##### 25 Man<sub>5</sub>GlcNAc<sub>2</sub>

The first step involves the selection or creation of a lower eukaryote that is able to produce a specific precursor structure of Man<sub>5</sub>GlcNAc<sub>2</sub>, which is able to accept *in vivo* GlcNAc by the action of a GlcNAc transferase I. This step requires the formation of a particular isomeric structure of 30 Man<sub>5</sub>GlcNAc<sub>2</sub>. This structure has to be formed within the cell at a high yield (in excess of 30%) since all subsequent manipulations are contingent on the

presence of this precursor.  $\text{Man}_5\text{GlcNAc}_2$  structures are necessary for complex N-glycan formation, however, their presence is by no means sufficient, since  $\text{Man}_5\text{GlcNAc}_2$  may occur in different isomeric forms, which may or may not serve as a substrate for GlcNAc transferase I. Most 5 glycosylation reactions are not complete and thus a particular protein generally contains a range of different carbohydrate structures (i.e. glycoforms) on its surface. The mere presence of trace amounts (less than 5%) of a particular structure like  $\text{Man}_5\text{GlcNAc}_2$  is of little practical relevance. It is the formation of a particular, GlcNAc transferase I accepting 10 intermediate (Structure I) in high yield (above 30%), which is required. The formation of this intermediate is necessary and subsequently allows for the *in vivo* synthesis of complex N-glycans.

One can select such lower eukaryotes from nature or alternatively genetically engineer existing fungi or other lower eukaryotes to provide the 15 structure *in vivo*. No lower eukaryote has been shown to provide such structures *in vivo* in excess of 1.8% of the total N-glycans (Maras et al., 1997), so a genetically engineered organism is preferred. Methods such as those described in U.S. Patent No. 5,595,900, may be used to identify the absence or presence of particular glycosyltransferases, mannosidases and 20 sugar nucleotide transporters in a target organism of interest.

Inactivation of Fungal Glycosylation Enzymes such as 1,2- $\alpha$ -mannosidase

The method described herein may be used to engineer the glycosylation pattern of a wide range of lower eukaryotes (e.g. *Hansenula polymorpha*, *Pichia stipitis*, *Pichia methanolica*, *Pichia* sp, *Kluyveromyces* sp, *Candida albicans*, *Aspergillus nidulans*, *Trichoderma reseei* etc.). *Pichia pastoris* is used to exemplify the required manipulation steps. Similar to other lower eukaryotes, *P. pastoris* processes  $\text{Man}_9\text{GlcNAc}_2$  structures in the ER with a 1,2- $\alpha$ -mannosidase to yield  $\text{Man}_8\text{GlcNAc}_2$ . Through the action 25 of several mannosyltransferases, this structure is then converted to hypermannosylated structures ( $\text{Man}_{>9}\text{GlcNAc}_2$ ), also known as mannans. In addition, it has been found that *P. pastoris* is able to add non-terminal 30

phosphate groups, through the action of mannosylphosphate transferases to the carbohydrate structure. This is contrary to the reactions found in mammalian cells, which involve the removal of mannose sugars as opposed to their addition. It is of particular importance to eliminate the ability of the 5 fungus to hypermannosylate the existing  $\text{Man}_2\text{GlcNAc}_2$  structure. This can be achieved by either selecting for a fungus that does not hypermannosylate, or by genetically engineering such a fungus.

Genes that are involved in this process have been identified in *Pichia pastoris* and by creating mutations in these genes one is able to reduce the 10 production of "undesirable" glycoforms. Such genes can be identified by homology to existing mannosyltransferases (e.g. OCH1, MNN4, MNN6, MNN1), found in other lower eukaryotes such as *C. albicans*, *Pichia angusta* or *S. cerevisiae* or by mutagenizing the host strain and selecting for a phenotype with reduced mannosylation. Based on homologies amongst 15 known mannosyltransferases and mannosylphosphate transferases, one may either design PCR primers, examples of which are shown in Table 2, or use genes or gene fragments encoding such enzymes as probes to identify homologues in DNA libraries of the target organism. Alternatively, one may be able to complement particular phenotypes in related organisms. For 20 example, in order to obtain the gene or genes encoding 1,6-mannosyltransferase activity in *P. pastoris*, one would carry out the following steps. OCH1 mutants of *S. cerevisiae* are temperature sensitive and are slow growers at elevated temperatures. One can thus identify functional homologues of OCH1 in *P. pastoris* by complementing an OCH1 mutant of 25 *S. cerevisiae* with a *P. pastoris* DNA or cDNA library. Such mutants of *S. cerevisiae* may be found at <http://genome-www.stanford.edu/Saccharomyces/> and are commercially available at <http://www.resgen.com/products/YEASTD.php3>. Mutants that display a normal growth phenotype at elevated temperature, after having been 30 transformed with a *P. pastoris* DNA library, are likely to carry an OCH1 homologue of *P. pastoris*. Such a library can be created by partially digesting chromosomal DNA of *P. pastoris* with a suitable restriction enzyme and after

inactivating the restriction enzyme ligating the digested DNA into a suitable vector, which has been digested with a compatible restriction enzyme. Suitable vectors are pRS314, a low copy (CEN6/ARS4) plasmid based on pBluescript containing the Trp1 marker (Sikorski, R. S., and Hieter, P., 1989, 5 Genetics 122, pg 19-27) or pFL44S, a high copy (2 $\mu$ ) plasmid based on a modified pUC19 containing the URA3 marker (Bonneaud, N., et al., 1991, Yeast 7, pg. 609-615). Such vectors are commonly used by academic researchers or similar vectors are available from a number of different vendors such as Invitrogen (Carlsbad, CA), Pharmacia (Piscataway, NJ), 10 New England Biolabs (Beverly, MA). Examples are pYES/GS, 2 $\mu$  origin of replication based yeast expression plasmid from Invitrogen, or Yep24 cloning vehicle from New England Biolabs. After ligation of the chromosomal DNA and the vector one may transform the DNA library into 15 strain of *S.cerevisiae* with a specific mutation and select for the correction of the corresponding phenotype. After sub-cloning and sequencing the DNA fragment that is able to restore the wild-type phenotype, one may use this fragment to eliminate the activity of the gene product encoded by OCH1 in 20 *P. pastoris*.

Alternatively, if the entire genomic sequence of a particular fungus of 25 interest is known, one may identify such genes simply by searching publicly available DNA databases, which are available from several sources such as NCBI, Swissprot etc. For example by searching a given genomic sequence or data base with a known 1,6 mannosyltransferase gene (OCH1) from *S. cerevisiae*, one can able to identify genes of high homology in such a genome, which a high degree of certainty encodes a gene that has 1,6 mannosyltransferase activity. Homologues to several known 30 mannosyltransferases from *S.cerevisiae* in *P.pastoris* have been identified using either one of these approaches. These genes have similar functions to genes involved in the mannosylation of proteins in *S. cerevisiae* and thus their deletion may be used to manipulate the glycosylation pattern in *P.pastoris* or any other fungus with similar glycosylation pathways.

The creation of gene knock-outs, once a given target gene sequence has been determined, is a well-established technique in the yeast and fungal molecular biology community, and can be carried out by anyone of ordinary skill in the art (R. Rothsteins, (1991) Methods in Enzymology, vol. 194, p. 5 281). In fact, the choice of a host organism may be influenced by the availability of good transformation and gene disruption techniques for such a host. If several mannosyltransferases have to be knocked out, the method developed by Alani and Kleckner allows for the repeated use of the URA3 markers to sequentially eliminate all undesirable endogenous 10 mannosyltransferase activity. This technique has been refined by others but basically involves the use of two repeated DNA sequences, flanking a counter selectable marker. For example: URA3 may be used as a marker to ensure the selection of a transformants that have integrated a construct. By flanking the URA3 marker with direct repeats one may first select for 15 transformants that have integrated the construct and have thus disrupted the target gene. After isolation of the transformants, and their characterization, one may counter select in a second round for those that are resistant to 5'FOA. Colonies that able to survive on plates containing 5'FOA have lost the URA3 marker again through a crossover event involving the repeats 20 mentioned earlier. This approach thus allows for the repeated use of the same marker and facilitates the disruption of multiple genes without requiring additional markers.

Eliminating specific mannosyltransferases, such as 1,6 25 mannosyltransferase (OCH1), mannosylphosphate transferases (MNN4, MNN6, or genes complementing *lbd* mutants) in *P. pastoris*, allows for the creation of engineered strains of this organism which synthesize primarily Man<sub>9</sub>GlcNAc<sub>2</sub> and thus can be used to further modify the glycosylation pattern to more closely resemble more complex human glycoform structures. A preferred embodiment of this method utilizes known DNA sequences, 30 encoding known biochemical glycosylation activities to eliminate similar or identical biochemical functions in *P. pastoris*, such that the glycosylation structure of the resulting genetically altered *P. pastoris* strain is modified.

Table 2.

<u>PCR primer A</u>	<u>PCR primer B</u>	<u>Target Gene(s) in <i>P. pastoris</i></u>	<u>Homologues</u>
ATGGCGAAGGCAGA TGGCAGT	TTAGTCCTTCCAAC TTCCTTC	1,6-mannosyltransferase	OCH1 <i>S.cerevisiae</i> , <i>Pichia albicans</i>
TAYTGGMGNGTNGA RCYNGAYATHAA	GCRTCNCCCCANCK YTCRTA	1,2-mannosyltransferase S	KTR/KRE family, <i>S.cerevisiae</i>

Legend: M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T, K = G or T, V = A or C or G, H = A or C or T, D = 5 A or G or T, B = C or G or T, N = G or A or T or C.

#### Incorporation of a Mannosidase into the Genetically Engineered Host

The process described herein enables one to obtain such a structure in 10 high yield for the purpose of modifying it to yield complex N-glycans. A successful scheme to obtain suitable  $\text{Man}_5\text{GlcNAc}_2$  structures must involve two parallel approaches: (1) reducing endogenous mannosyltransferase activity and (2) removing 1,2- $\alpha$ -mannose by mannosidases to yield high levels of suitable  $\text{Man}_5\text{GlcNAc}_2$  structures. What distinguishes this method 15 from the prior art is that it deals directly with those two issues. As the work of Chiba and coworkers demonstrates, one can reduce  $\text{Man}_8\text{GlcNAc}_2$  structures to a  $\text{Man}_5\text{GlcNAc}_2$  isomer in *S. cerevisiae*, by engineering the presence of a fungal mannosidase from *A. saitoi* into the ER. The shortcomings of their approach are twofold: (1) insufficient amounts of 20  $\text{Man}_5\text{GlcNAc}_2$  are formed in the extra-cellular glycoprotein fraction (10%) and (2) it is not clear that the *in vivo* formed  $\text{Man}_5\text{GlcNAc}_2$  structure in fact is able to accept GlcNAc by action of GlcNAc transferase I. If several glycosylation sites are present in a desired protein the probability (P) of 25 obtaining such a protein in a correct form follows the relationship  $P=(F)^n$ , where n equals the number of glycosylation sites, and F equals the fraction of desired glycoforms. A glycoprotein with three glycosylation sites would

have a 0.1% chance of providing the appropriate precursors for complex and hybrid N-glycan processing on all of its glycosylation sites, which limits the commercial value of such an approach.

Most enzymes that are active in the ER and Golgi apparatus of

5 *S.cerevisiae* have pH optima that are between 6.5 and 7.5 (see Table 3). All previous approaches to reduce mannosylation by the action of recombinant mannosidases have concentrated on enzymes that have a pH optimum around pH 5.0 (Martinet et al., 1998, and Chiba et al., 1998), even though the activity of these enzymes is reduced to less than 10% at pH 7.0 and thus

10 most likely provide insufficient activity at their point of use, the ER and early Golgi of *P.pastoris* and *S.cerevisiae*. A preferred process utilizes an  $\alpha$ -mannosidase *in vivo*, where the pH optimum of the mannosidase is within 1.4 pH units of the average pH optimum of other representative marker enzymes localized in the same organelle(s). The pH optimum of the enzyme

15 to be targeted to a specific organelle should be matched with the pH optimum of other enzymes found in the same organelle, such that the maximum activity per unit enzyme is obtained. Table 3 summarizes the activity of mannosidases from various sources and their respective pH optima. Table 4 summarizes their location.

Table 3. Mannosidases and their pH optimum.

Source	Enzyme	pH optimum	Reference
<i>Aspergillus saitoi</i>	1,2- $\alpha$ - mannosidase	5.0	Ichishima et al., 1999 Biochem. J. 339(Pt 3):589-597
<i>Trichoderma reesei</i>	1,2- $\alpha$ - mannosidase	5.0	Maras et al., 2000 J. Biotechnol. 77(2-3):255-263
<i>Penicillium citrinum</i>	1,2- $\alpha$ -D-mannosidase	5.0	Yoshida et al., 1993 Biochem. J. 290(Pt 2):349-354
<i>Aspergillus nidulans</i>	1,2- $\alpha$ - mannosidase	6.0	Eades and Hintz, 2000
<i>Homo sapiens</i> IA(Golgi)	1,2- $\alpha$ - mannosidase	6.0	
<i>Homo sapiens</i> IB (Golgi)	1,2- $\alpha$ - mannosidase	6.0	
<i>Lepidopteran</i> insect cells	Type I 1,2- $\alpha$ -Man <sub>6</sub> -mannosidase	6.0	Ren et al., 1995 Biochem. 34(8):2489-2495
<i>Homo sapiens</i>	$\alpha$ - D-mannosidase	6.0	Chandrasekaran et al., 1984 Cancer Res. 44(9):4059-68
<i>Xanthomonas</i> <i>manihotis</i>	1,2,3- $\alpha$ - mannosidase	6.0	
<i>Mouse</i> IB (Golgi)	1,2- $\alpha$ - mannosidase	6.5	Schneikert and Herscovics, 1994 Glycobiology. 4(4):445-50
<i>Bacillus</i> sp. (secreted)	1,2- $\alpha$ -D-mannosidase	7.0	Maruyama et al., 1994 Carbohydrate Res. 251:89-98

When one attempts to trim high mannose structures to yield  $\text{Man}_5\text{GlcNAc}_2$  in the ER or the Golgi apparatus of *S. cerevisiae*, one may choose any enzyme or combination of enzymes that (1) has/have a sufficiently close pH optimum (i.e. between pH 5.2 and pH 7.8), and (2)

5 is/are known to generate, alone or in concert, the specific isomeric  $\text{Man}_5\text{GlcNAc}_2$  structure required to accept subsequent addition of GlcNAc by GnT I. Any enzyme or combination of enzymes that has/have shown to generate a structure that can be converted to  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  by GnT I *in vitro* would constitute an appropriate choice. This knowledge may be

10 obtained from the scientific literature or experimentally by determining that a potential mannosidase can convert  $\text{Man}_8\text{GlcNAc}_2\text{-PA}$  to  $\text{Man}_5\text{GlcNAc}_2\text{-PA}$  and then testing, if the obtained  $\text{Man}_5\text{GlcNAc}_2\text{-PA}$  structure can serve a substrate for GnT I and UDP-GlcNAc to give  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  *in vitro*. For example, mannosidase IA from a human or murine source would be an

15 appropriate choice.

*1,2-Mannosidase Activity in the ER and Golgi*

Previous approaches to reduce mannosylation by the action of cloned exogenous mannosidases have failed to yield glycoproteins having a sufficient fraction (e.g. >27 mole %) of *N*-glycans having the structure

20  $\text{Man}_5\text{GlcNAc}_2$  (Martinet et al., 1998, and Chiba et al., 1998). These enzymes should function efficiently in ER or Golgi apparatus to be effective in converting nascent glycoproteins. Whereas the two mannosidases utilized in the prior art (from *A. saitoi* and *T. reesei*) have pH optima of 5.0, most enzymes that are active in the ER and Golgi apparatus of yeast (e.g. *S. cerevisiae*) have pH optima that are between 6.5 and 7.5 (see Table 3). Since the glycosylation of proteins is a highly evolved and efficient process, it can be concluded that the internal pH of the ER and the Golgi is also in the range of about 6-8. At pH 7.0, the activity of the mannosidases used in the prior art is reduced to less than 10%, which is insufficient for the efficient

25 production of  $\text{Man}_5\text{GlcNAc}_2$  *in vivo*.

Table 4. Cellular location and pH optima of various glycosylation-related enzymes of *S. cerevisiae*.

Gene	Activity	Location	pH optimum	Author(s)
<i>Ktr1</i>	$\alpha$ - 1,2 mannosyltransferase	Golgi	7.0	Romero et al., 1997 Biochem. J. 321(Pt 2):289-295
<i>Mns1</i>	$\alpha$ - 1,2- mannosidase	ER	6.5	
<i>CWH41</i>	glucosidase I	ER	6.8	
--	mannosyltransferase	Golgi	7-8	Lehele and Tanner, 1974 Biochim. Biophys. Acta 350(1):225-235
<i>Kre2</i>	$\alpha$ - 1,2 mannosyltransferase	Golgi	6.5-9.0	Romero et al., 1997

The  $\alpha$ -1,2-mannosidase enzyme should have optimal activity at a pH between 5.1 and 8.0. In a preferred embodiment, the enzyme has an optimal activity at a pH between 5.9 and 7.5. The optimal pH may be determined under *in vitro* assay conditions. Preferred mannosidases include those listed in Table 3 having appropriate pH optima, e.g. *Aspergillus nidulans*, *Homo sapiens* IA(Golgi), *Homo sapiens* IB (Golgi), *Lepidopteran* insect cells (IPLB-SF21AE), *Homo sapiens*, mouse IB (Golgi), and *Xanthomonas manihotis*. In a preferred embodiment, a single cloned mannosidase gene is expressed in the host organism. However, in some cases it may be desirable to express several different mannosidase genes, or several copies of one particular gene, in order to achieve adequate production of  $\text{Man}_5\text{GlcNAc}_2$ . In cases where multiple genes are used, the encoded mannosidases should all have pH optima within the preferred range of 5.1 to 8.0, or especially between 5.9 and 7.5. In an especially preferred embodiment mannosidase

activity is targeted to the ER or cis Golgi, where the early reactions of glycosylation occur.

*Formation of complex N-glycans*

A second step of the process involves the sequential addition of

5 sugars to the nascent carbohydrate structure by engineering the expression of glucosyltransferases into the Golgi apparatus. This process first requires the functional expression of GnT I in the early or medial Golgi apparatus as well as ensuring the sufficient supply of UDP-GlcNAc.

**Integration Sites**

10 Since the ultimate goal of this genetic engineering effort is a robust protein production strain that is able to perform well in an industrial fermentation process, the integration of multiple genes into the fungal chromosome involves careful planning. The engineered strain will most likely have to be transformed with a range of different genes, and these genes will

15 have to be transformed in a stable fashion to ensure that the desired activity is maintained throughout the fermentation process. Any combination of the following enzyme activities will have to be engineered into the fungal protein expression host: sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, glucosyltransferases, GlcNAc

20 transferases, ER and Golgi specific transporters (e.g. sym and antiport transporters for UDP-galactose and other precursors), other enzymes involved in the processing of oligosaccharides, and enzymes involved in the synthesis of activated oligosaccharide precursors such as UDP-galactose, CMP-N-acetylneuraminic acid. At the same time a number of genes which

25 encode enzymes known to be characteristic of non-human glycosylation reactions, will have to be deleted.

**Targeting of glycosyltransferases to specific organelles:**

Glycosyltransferases and mannosidases line the inner (luminal) surface of the ER and Golgi apparatus and thereby provide a "catalytic" 30 surface that allows for the sequential processing of glycoproteins as they proceed through the ER and Golgi network. In fact the multiple compartments of the cis, medial, and trans Golgi and the trans-Golgi

Network (TGN), provide the different localities in which the ordered sequence of glycosylation reactions can take place. As a glycoprotein proceeds from synthesis in the ER to full maturation in the late Golgi or TGN, it is sequentially exposed to different glycosidases, mannosidases and 5 glycosyltransferases such that a specific carbohydrate structure may be synthesized. Much work has been dedicated to revealing the exact mechanism by which these enzymes are retained and anchored to their respective organelle. The evolving picture is complex but evidence suggests that stem region, membrane spanning region and cytoplasmic tail 10 individually or in concert direct enzymes to the membrane of individual organelles and thereby localize the associated catalytic domain to that locus.

Targeting sequences are well known and described in the scientific literature and public databases, as discussed in more detail below with respect to libraries for selection of targeting sequences and targeted enzymes.

15 Method for Producing a Library to Produce Modified Glycosylation Pathways

A library including at least two genes encoding exogenous glycosylation enzymes is transformed into the host organism, producing a genetically mixed population. Transformants having the desired 20 glycosylation phenotypes are then selected from the mixed population. In a preferred embodiment, the host organism is a yeast, especially *P. pastoris*, and the host glycosylation pathway is modified by the operative expression of one or more human or animal glycosylation enzymes, yielding protein *N*-glycans similar or identical to human glycoforms. In an especially preferred 25 embodiment, the DNA library includes genetic constructs encoding fusions of glycosylation enzymes with targeting sequences for various cellular loci involved in glycosylation especially the ER, cis Golgi, medial Golgi, or trans Golgi.

Examples of modifications to glycosylation which can be effected 30 using method are: (1) engineering an eukaryotic microorganism to trim mannose residues from  $\text{Man}_8\text{GlcNAc}_2$  to yield  $\text{Man}_5\text{GlcNAc}_2$  as a protein *N*-glycan; (2) engineering an eukaryotic microorganism to add an

*N*-acetylglucosamine (GlcNAc) residue to Man<sub>5</sub>GlcNAc<sub>2</sub> by action of GlcNAc transferase I; (3) engineering an eukaryotic microorganism to functionally express an enzyme such as an *N*-acetylglucosamine transferase (GnT I, GnT II, GnT III, GnT IV, GnT V, GnT VI), mannosidase II, 5 fucosyltransferase, galactosyl transferase (GalT) or sialyltransferases (ST).

By repeating the method, increasingly complex glycosylation pathways can be engineered into the target microorganism. In one preferred embodiment, the host organism is transformed two or more times with DNA libraries including sequences encoding glycosylation activities. Selection of 10 desired phenotypes may be performed after each round of transformation or alternatively after several transformations have occurred. Complex glycosylation pathways can be rapidly engineered in this manner.

#### *DNA Libraries*

It is necessary to assemble a DNA library including at least two 15 exogenous genes encoding glycosylation enzymes. In addition to the open reading frame sequences, it is generally preferable to provide each library construct with such promoters, transcription terminators, enhancers, ribosome binding sites, and other functional sequences as may be necessary to ensure effective transcription and translation of the genes upon 20 transformation into the host organism. Where the host is *Pichia pastoris*, suitable promoters include, for example, the *AOX1*, *AOX2*, *DAS*, and *P40* promoters. It is also preferable to provide each construct with at least one selectable marker, such as a gene to impart drug resistance or to complement a host metabolic lesion. The presence of the marker is useful in the 25 subsequent selection of transformants; for example, in yeast the *URA3*, *HIS4*, *SUC2*, *G418*, *BLA*, or *SH BLE* genes may be used.

In some cases the library may be assembled directly from existing or 30 wild-type genes. In a preferred embodiment however the DNA library is assembled from the fusion of two or more sub-libraries. By the in-frame ligation of the sub-libraries, it is possible to create a large number of novel genetic constructs encoding useful targeted glycosylation activities. For example, one useful sub-library includes DNA sequences encoding any

combination of enzymes such as sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, glucosyltransferases, and GlcNAc transferases. Preferably, the enzymes are of human origin, although other mammalian, animal, or fungal enzymes are also useful. In a preferred 5 embodiment, genes are truncated to give fragments encoding the catalytic domains of the enzymes. By removing endogenous targeting sequences, the enzymes may then be redirected and expressed in other cellular loci. The choice of such catalytic domains may be guided by the knowledge of the particular environment in which the catalytic domain is subsequently to be 10 active. For example, if a particular glycosylation enzyme is to be active in the late Golgi, and all known enzymes of the host organism in the late Golgi have a certain pH optimum, then a catalytic domain is chosen which exhibits adequate activity at that pH.

Another useful sub-library includes DNA sequences encoding signal 15 peptides that result in localization of a protein to a particular location within the ER, Golgi, or trans Golgi network. These signal sequences may be selected from the host organism as well as from other related or unrelated organisms. Membrane-bound proteins of the ER or Golgi typically may include, for example, N-terminal sequences encoding a cytosolic tail (ct), a 20 transmembrane domain (tmd), and a stem region (sr). The ct, tmd, and sr sequences are sufficient individually or in combination to anchor proteins to the inner (luminal) membrane of the organelle. Accordingly, a preferred embodiment of the sub-library of signal sequences includes ct, tmd, and/or sr sequences from these proteins. In some cases it is desirable to provide the 25 sub-library with varying lengths of sr sequence. This may be accomplished by PCR using primers that bind to the 5' end of the DNA encoding the cytosolic region and employing a series of opposing primers that bind to various parts of the stem region. Still other useful sources of signal sequences include retrieval signal peptides, e.g. the tetrapeptides HDEL or 30 KDEL, which are typically found at the C-terminus of proteins that are transported retrograde into the ER or Golgi. Still other sources of signal sequences include (a) type II membrane proteins, (b) the enzymes listed in

Table 3, (c) membrane spanning nucleotide sugar transporters that are localized in the Golgi, and (d) sequences referenced in Table 5.

Table 5. Sources of useful compartmental targeting sequences

<b>Gene or Sequence</b>	<b>Organism</b>	<b>Function</b>	<b>Location of Gene Product</b>
<i>MnsI</i>	<i>S. cerevisiae</i>	$\alpha$ -1,2-mannosidase	ER
<i>OCH1</i>	<i>S. cerevisiae</i>	1,6-mannosyltransferase	Golgi (cis)
<i>MNN2</i>	<i>S. cerevisiae</i>	1,2-mannosyltransferase	Golgi (medial)
<i>MNN1</i>	<i>S. cerevisiae</i>	1,3-mannosyltransferase	Golgi (trans)
<i>OCH1</i>	<i>P. pastoris</i>	1,6-mannosyltransferase	Golgi (cis)
2,6 ST	<i>H. sapiens</i>	2,6-sialyltransferase	trans Golgi network
UDP-Gal T	<i>S. pombe</i>	UDP-Gal transporter	Golgi
<i>Mnt1</i>	<i>S. cerevisiae</i>	1,2-mannosyltransferase	Golgi (cis)
HDEL at C-terminus	<i>S. cerevisiae</i>	retrieval signal	ER

In any case, it is highly preferred that signal sequences are selected which are appropriate for the enzymatic activity or activities which are to be engineered into the host. For example, in developing a modified microorganism capable of terminal sialylation of nascent *N*-glycans, a process which occurs in the late Golgi in humans, it is desirable to utilize a sub-library of signal sequences derived from late Golgi proteins. Similarly, the trimming of  $\text{Man}_8\text{GlcNAc}_2$  by an  $\alpha$ -1,2-mannosidase to give  $\text{Man}_5\text{GlcNAc}_2$  is an early step in complex *N*-glycan formation in humans. It is therefore desirable to have this reaction occur in the ER or early Golgi of

an engineered host microorganism. A sub-library encoding ER and early Golgi retention signals is used.

In a preferred embodiment, a DNA library is then constructed by the in-frame ligation of a sub-library including DNA encoding signal sequences 5 with a sub-library including DNA encoding glycosylation enzymes or catalytically active fragments thereof. The resulting library includes synthetic genes encoding fusion proteins. In some cases it is desirable to provide a signal sequence at the N-terminus of a fusion protein, or in other cases at the C-terminus. In some cases signal sequences may be inserted 10 within the open reading frame of an enzyme, provided the protein structure of individual folded domains is not disrupted.

The method is most effective when a DNA library transformed into the host contains a large diversity of sequences, thereby increasing the probability that at least one transformant will exhibit the desired phenotype. 15 Accordingly, prior to transformation, a DNA library or a constituent sub-library may be subjected to one or more rounds of gene shuffling, error prone PCR, or *in vitro* mutagenesis.

#### *Transformation*

The DNA library is then transformed into the host organism. In 20 yeast, any convenient method of DNA transfer may be used, such as electroporation, the lithium chloride method, or the spheroplast method. To produce a stable strain suitable for high-density fermentation, it is desirable to integrate the DNA library constructs into the host chromosome. In a preferred embodiment, integration occurs via homologous recombination, 25 using techniques known in the art. For example, DNA library elements are provided with flanking sequences homologous to sequences of the host organism. In this manner integration occurs at a defined site in the host genome, without disruption of desirable or essential genes. In an especially preferred embodiment, library DNA is integrated into the site of an undesired 30 gene in a host chromosome, effecting the disruption or deletion of the gene. For example, integration into the sites of the *OCH1*, *MNN1*, or *MNN4* genes allows the expression of the desired library DNA while preventing the

expression of enzymes involved in yeast hypermannosylation of glycoproteins. In other embodiments, library DNA may be introduced into the host via a chromosome, plasmid, retroviral vector, or random integration into the host genome. In any case, it is generally desirable to include with 5 each library DNA construct at least one selectable marker gene to allow ready selection of host organisms that have been stably transformed. Recyclable marker genes such as *ura3*, which can be selected for or against, are especially suitable.

*Selection Process*

10 After transformation of the host strain with the DNA library, transformants displaying the desired glycosylation phenotype are selected. Selection may be performed in a single step or by a series of phenotypic enrichment and/or depletion steps using any of a variety of assays or detection methods. Phenotypic characterization may be carried out manually 15 or using automated high-throughput screening equipment. Commonly a host microorganism displays protein *N*-glycans on the cell surface, where various glycoproteins are localized. Accordingly intact cells may be screened for a desired glycosylation phenotype by exposing the cells to a lectin or antibody that binds specifically to the desired *N*-glycan. A wide variety of 20 oligosaccharide-specific lectins are available commercially (EY Laboratories, San Mateo, CA). Alternatively, antibodies to specific human or animal *N*-glycans are available commercially or may be produced using standard techniques. An appropriate lectin or antibody may be conjugated to a reporter molecule, such as a chromophore, fluorophore, radioisotope, or an 25 enzyme having a chromogenic substrate (Guillen et al., 1998. *Proc. Natl. Acad. Sci. USA* 95(14): 7888-7892). Screening may then be performed using analytical methods such as spectrophotometry, fluorimetry, fluorescence activated cell sorting, or scintillation counting. In other cases, it may be necessary to analyze isolated glycoproteins or *N*-glycans from 30 transformed cells. Protein isolation may be carried out by techniques known in the art. In cases where an isolated *N*-glycan is required, an enzyme such as endo- $\beta$ -*N*-acetylglucosaminidase (Genzyme Co., Boston, MA) may be

used to cleave the *N*-glycans from glycoproteins. Isolated proteins or *N*-glycans may then be analyzed by liquid chromatography (e.g. HPLC), mass spectroscopy, or other suitable means. U.S. Patent No. 5,595,900 teaches several methods by which cells with desired extracellular

5 carbohydrate structures may be identified. Prior to selection of a desired transformant, it may be desirable to deplete the transformed population of cells having undesired phenotypes. For example, when the method is used to engineer a functional mannosidase activity into cells, the desired transformants will have lower levels of mannose in cellular glycoprotein.

10 Exposing the transformed population to a lethal radioisotope of mannose in the medium depletes the population of transformants having the undesired phenotype, i.e. high levels of incorporated mannose. Alternatively, a cytotoxic lectin or antibody, directed against an undesirable *N*-glycan, may be used to deplete a transformed population of undesired phenotypes.

15 Methods for Providing Sugar Nucleotide Precursors to the Golgi Apparatus

For a glycosyltransferase to function satisfactorily in the Golgi, it is necessary for the enzyme to be provided with a sufficient concentration of an appropriate nucleotide sugar, which is the high-energy donor of the sugar moiety added to a nascent glycoprotein. These nucleotide sugars to the

20 appropriate compartments are provided by expressing an exogenous gene encoding a sugar nucleotide transporter in the host microorganism. The choice of transporter enzyme is influenced by the nature of the exogenous glycosyltransferase being used. For example, a GlcNAc transferase may require a UDP-GlcNAc transporter, a fucosyltransferase may require a GDP-

25 fucose transporter, a galactosyltransferase may require a UDP-galactose transporter, or a sialyltransferase may require a CMP-sialic acid transporter.

The added transporter protein conveys a nucleotide sugar from the cytosol into the Golgi apparatus, where the nucleotide sugar may be reacted by the glycosyltransferase, e.g. to elongate an *N*-glycan. The reaction

30 liberates a nucleoside diphosphate or monophosphate, e.g. UDP, GDP, or CMP. As accumulation of a nucleoside diphosphate inhibits the further activity of a glycosyltransferase, it is frequently also desirable to provide an

expressed copy of a gene encoding a nucleotide diphosphatase. The diphosphatase (specific for UDP or GDP as appropriate) hydrolyzes the diphosphonucleoside to yield a nucleoside monophosphate and inorganic phosphate. The nucleoside monophosphate does not inhibit the 5 glycotransferase and in any case is exported from the Golgi by an endogenous cellular system. Suitable transporter enzymes, which are typically of mammalian origin, are described below.

*Examples*

The use of the above general method may be understood by reference 10 to the following non-limiting examples. Examples of preferred embodiments are also summarized in Table 6.

**Example 1: Engineering of *P. pastoris* with  $\alpha$ -1,2-Mannosidase to produce insulin.**

An  $\alpha$ -1,2-mannosidase is required for the trimming of  $\text{Man}_9\text{GlcNAc}_2$  15 to yield  $\text{Man}_5\text{GlcNAc}_2$ , an essential intermediate for complex *N*-glycan formation. An *OCHI* mutant of *P. pastoris* is engineered to express secreted human interferon- $\alpha$  under the control of an *aox* promoter. A DNA library is constructed by the in-frame ligation of the catalytic domain of human mannosidase IB (an  $\alpha$ -1,2-mannosidase) with a sub-library including 20 sequences encoding early Golgi localization peptides. The DNA library is then transformed into the host organism, resulting in a genetically mixed population wherein individual transformants each express interferon- $\beta$  as well as a synthetic mannosidase gene from the library. Individual transformant colonies are cultured and the production of interferon is 25 induced by addition of methanol. Under these conditions, over 90% of the secreted protein includes interferon- $\beta$ . Supernatants are purified to remove salts and low-molecular weight contaminants by  $\text{C}_{18}$  silica reversed-phase chromatography. Desired transformants expressing appropriately targeted, active  $\alpha$ -1,2-mannosidase produce interferon- $\beta$  including *N*-glycans of the 30 structure  $\text{Man}_5\text{GlcNAc}_2$ , which has a reduced molecular mass compared to the interferon of the parent strain. The purified supernatants including

interferon- $\beta$  are analyzed by MALDI-TOF mass spectroscopy and colonies expressing the desired form of interferon- $\beta$  are identified.

**Example 2: Engineering of Strain to express GlcNAc Transferase I**

GlcNAc Transferase I activity is required for the maturation of complex *N*-glycans.  $\text{Man}_5\text{GlcNAc}_2$  may only be trimmed by mannosidase II, a necessary step in the formation of human glycoforms, after the addition of GlcNAc to the terminal  $\alpha$ -1,3 mannose residue by GlcNAc Transferase I (Schachter, 1991 *Glycobiology* 1(5):453-461). Accordingly a library is prepared including DNA fragments encoding suitably targeted GlcNAc Transferase I genes. The host organism is a strain, e.g. a yeast, that is deficient in hypermannosylation (e.g. an *OCH1* mutant), provides the substrate UDP-GlcNAc in the Golgi and/or ER, and provides *N*-glycans of the structure  $\text{Man}_5\text{GlcNAc}_2$  in the Golgi and/or ER. After transformation of the host with the DNA library, the transformants are screened for those having the highest concentration of terminal GlcNAc on the cell surface, or alternatively secrete the protein having the highest terminal GlcNAc content. Such a screen is performed using a visual method (e.g. a staining procedure), a specific terminal GlcNAc binding antibody, or a lectin. Alternatively the desired transformants exhibit reduced binding of certain lectins specific for terminal mannose residues.

**Example 3: Engineering of Strains with a Mannosidase II**

In another example, it is desirable in order to generate a human glycoform in a microorganism to remove the two remaining terminal mannoses from the structure  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  by action of a mannosidase II. A DNA library including sequences encoding cis and medial Golgi localization signals is fused in-frame to a library encoding mannosidase II catalytic domains. The host organism is a strain, e.g. a yeast, that is deficient in hypermannosylation (e.g. an *OCH1* mutant) and provides *N*-glycans having the structure  $\text{GlcNAcMan}_3\text{GlcNAc}_2$  in the Golgi and/or ER. After transformation, organisms having the desired glycosylation phenotype are selected. An *in vitro* assay is used in one method. The desired structure  $\text{GlcNAcMan}_3\text{GlcNAc}_2$  (but not the undesired

GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>) is a substrate for the enzyme GlcNAc Transferase II. Accordingly, single colonies may be assayed using this enzyme *in vitro* in the presence of the substrate, UDP-GlcNAc. The release of UDP is determined either by HPLC or an enzymatic assay for UDP. Alternatively 5 radioactively labeled UDP-GlcNAc is used.

The foregoing *in vitro* assays are conveniently performed on individual colonies using high-throughput screening equipment. Alternatively a lectin binding assay is used. In this case the reduced binding of lectins specific for terminal mannoses allows the selection of 10 transformants having the desired phenotype. For example, *Galanthus nivalis* lectin binds specifically to terminal  $\alpha$ -1,3-mannose, the concentration of which is reduced in the presence of operatively expressed mannosidase II activity. In one suitable method, *G. nivalis* lectin attached to a solid agarose support (available from Sigma Chemical, St. Louis, MO) is used to deplete 15 the transformed population of cells having high levels of terminal  $\alpha$ -1,3-mannose.

**Example 4: Engineering of organisms to express Sialyltransferase**

The enzymes  $\alpha$ 2,3-sialyltransferase and  $\alpha$ 2,6-sialyltransferase add terminal sialic acid to galactose residues in nascent human *N*-glycans, 20 leading to mature glycoproteins. In human the reactions occur in the trans Golgi or TGN. Accordingly a DNA library is constructed by the in-frame fusion of sequences encoding sialyltransferase catalytic domains with sequences encoding trans Golgi or TGN localization signals. The host organism is a strain, e.g. a yeast, that is deficient in hypermannosylation (e.g. 25 an *OCH1* mutant), which provides *N*-glycans having terminal galactose residues in the trans Golgi or TGN, and provides a sufficient concentration of CMP-sialic acid in the trans Golgi or TGN. Following transformation, transformants having the desired phenotype are selected using a fluorescent antibody specific for *N*-glycans having a terminal sialic acid.

**Example 5: Method of engineering strains to express UDP-GlcNAc Transporter**

The cDNA of human Golgi UDP-GlcNAc transporter has been cloned by Ishida and coworkers. (Ishida, N., et al. 1999 *J. Biochem.* 126(1): 68-77. Guillen and coworkers have cloned the canine kidney Golgi UDP-GlcNAc transporter by phenotypic correction of a *Kluyveromyces lactis* mutant deficient in Golgi UDP-GlcNAc transport. (Guillen, E., et al. 1998). Thus a mammalian Golgi UDP-GlcNAc transporter gene has all of the necessary information for the protein to be expressed and targeted functionally to the Golgi apparatus of yeast.

**Example 6: Method of engineering strains to express GDP-Fucose Transporter.**

The rat liver Golgi membrane GDP-fucose transporter has been identified and purified by Puglielli, L. and C. B. Hirschberg 1999 *J. Biol. Chem.* 274(50):35596-35600. The corresponding gene can be identified using standard techniques, such as *N*-terminal sequencing and Southern blotting using a degenerate DNA probe. The intact gene can then be expressed in a host microorganism that also expresses a fucosyltransferase.

**Example 7: Method of engineering strains to express UDP-Galactose Transporter**

Human UDP-galactose (UDP-Gal) transporter has been cloned and shown to be active in *S. cerevisiae*. (Kainuma, M., et al. 1999 *Glycobiology* 9(2): 133-141). A second human UDP-galactose transporter (hUGT1) has been cloned and functionally expressed in Chinese Hamster Ovary Cells. Aoki, K., et al. 1999 *J. Biochem.* 126(5): 940-950. Likewise Segawa and coworkers have cloned a UDP-galactose transporter from *Schizosaccharomyces pombe* (Segawa, H., et al. 1999 *Febs Letters* 451(3): 295-298).

*CMP-Sialic Acid Transporter*

Human CMP-sialic acid transporter (hCST) has been cloned and expressed in Lec 8 CHO cells by Aoki and coworkers (1999). Molecular cloning of the hamster CMP-sialic acid transporter has also been achieved

(Eckhardt and Gerardy Schahn 1997 *Eur. J. Biochem.* 248(1): 187-192). The functional expression of the murine CMP-sialic acid transporter was achieved in *Saccharomyces cerevisiae* by Berninsone, P., et al. 1997 *J. Biol. Chem.* 272(19):12616-12619.

5 Table6. Examples of preferred embodiments of the methods for modifying glycosylation in a eukaryotic microorganism, e.g. *Pichia pastoris*

Desired Structure	Suitable Catalytic Activities	Suitable Sources of Localization Sequences	Suitable Gene Deletions	Suitable Transporters and/or Phosphatases
Man <sub>5</sub> GlcNAc <sub>2</sub>	α-1,2-mannosidase (murine, human, <i>Bacillus</i> sp., <i>A. nidulans</i> )	Mns1 (N-terminus, <i>S. cerevisiae</i> ) Och1 (N-terminus, <i>S. cerevisiae</i> , <i>P. pastoris</i> ) Ktr1 Mnn9 Mnt1 ( <i>S. cerevisiae</i> ) KDEL, HDEL (C-terminus)	OCH1 MNN4 MNN6	none
GlcNAcMan <sub>5</sub> GlcNAc <sub>2</sub>	GlcNAc Transferase I, (human, murine, rat etc.)	Och1 (N-terminus, <i>S. cerevisiae</i> , <i>P. pastoris</i> ) KTR1 (N-terminus) KDEL, HDEL (C-terminus) Mnn1 (N-terminus, <i>S. cerevisiae</i> ) Mnt1 (N-terminus, <i>S. cerevisiae</i> ) GDase (N-terminus, <i>S. cerevisiae</i> )	OCH1 MNN4 MNN6	UDP-GlcNAc transporter (human, murine, <i>K. lactis</i> ) UDPase (human)

GlcNAcMan <sub>3</sub> GlcNAc <sub>2</sub>	mannosidase II	Ktr1 Mnn1 (N-terminus, <i>S. cerevisiae</i> ) Mnt1(N-terminus, <i>S. cerevisiae</i> ) Kre2/Mnt1 ( <i>S. cerevisiae</i> ) Kre2 ( <i>P. pastoris</i> ) Ktr1 ( <i>S. cerevisiae</i> ) Ktr1 ( <i>P. pastoris</i> ) Mnn1 ( <i>S. cerevisiae</i> )	OCH1 MNN4 MNN6	UDP-GlcNAc transporter (human, murine, <i>K. lactis</i> ) UDPase (human)
GlcNAc <sub>(2-4)</sub> Man <sub>3</sub> GlcNAc <sub>2</sub>	GlcNAc Transferase II, III, IV, V (human, murine)	Mnn1 (N-terminus, <i>S. cerevisiae</i> ) Mnt1 (N-terminus, <i>S. cerevisiae</i> ) Kre2/Mnt1 ( <i>S. cerevisiae</i> ) Kre2 ( <i>P. pastoris</i> ) Ktr1 ( <i>S. cerevisiae</i> ) Ktr1 ( <i>P. pastoris</i> ) Mnn1 ( <i>S. cerevisiae</i> )	OCH1 MNN4 MNN6	UDP-GlcNAc transporter (human, murine, <i>K. lactis</i> ) UDPase (human)
Gal <sub>(1-4)</sub> GlcNAc <sub>(2-4)</sub> -Man <sub>3</sub> GlcNAc <sub>2</sub>	β-1,4-Galactosyl transferase (human)	Mnn1 (N-terminus, <i>S. cerevisiae</i> ) Mnt1(N-terminus, <i>S. cerevisiae</i> ) Kre2/Mnt1 ( <i>S. cerevisiae</i> ) Kre2 ( <i>P. pastoris</i> ) Ktr1 ( <i>S. cerevisiae</i> ) Ktr1 ( <i>P. pastoris</i> ) Mnn1 ( <i>S. cerevisiae</i> )	OCH1 MNN4 MNN6	UDP-Galactose transporter (human, <i>S. pombe</i> )
NANA <sub>(1-4)</sub> -Gal <sub>(1-4)</sub> GlcNAc <sub>(2-4)</sub> -Man <sub>3</sub> GlcNAc <sub>2</sub>	α-2,6-Sialyltransferase (human) α-2,3-Sialyltransferase	KTR1 MNN1 (N-terminus, <i>S. cerevisiae</i> ) MNT1 (N-terminus, <i>S. cerevisiae</i> ) Kre2/Mnt1 ( <i>S. cerevisiae</i> ) Kre2 ( <i>P. pastoris</i> ) Ktr1 ( <i>S. cerevisiae</i> ) Ktr1 ( <i>P. pastoris</i> ) MNN1 ( <i>S. cerevisiae</i> )	OCH1 MNN4 MNN6	CMP-Sialic acid transporter (human)

**Table 7: DNA and Protein Sequence Resources**

1. European Bioinformatics Institute (EBI) is a centre for research and services in bioinformatics: <http://www.ebi.ac.uk/>
2. Swissprot database: <http://www.expasy.ch/spr>
- 5 3. List of known glycosyltransferases and their origin.  
 $\beta$ 1,2 (GnT I) EC 2.4.1.101
4. human cDNA, Kumar et al (1990) Proc. Natl. Acad. Sci. USA 87:9948-9952
5. human gene, Hull et al (1991) Biochem. Biophys. Res. Commun. 10 176:608-615
6. mouse cDNA, Kumar et al (1992) Glycobiology 2:383-393
7. mouse gene, Pownall et al (1992) Genomics 12:699-704
8. murine gene (5' flanking, non-coding), Yang et al (1994) Glycobiology 5:703-712
- 15 9. rabbit cDNA, Sarkar et al (1991) Proc. Natl. Acad. Sci. USA 88:234-238
10. rat cDNA, Fukada et al (1994) Biosci.Biotechnol.Biochem. 58:200-201  
1,2 (GnT II) EC 2.4.1.143
11. human gene, Tan et al (1995) Eur. J. Biochem. 231:317-328
12. rat cDNA, D'Agostaro et al (1995) J. Biol. Chem. 270:15211-15221
13.  $\beta$ 1,4 (GnT III) EC 2.4.1.144
14. human cDNA, Ihara et al (1993) J. Biochem. 113:692-698
15. murine gene, Bhaumik et al (1995) Gene 164:295-300
- 25 16. rat cDNA, Nishikawa et al (1992) J. Biol. Chem. 267:18199-18204  
 $\beta$ 1,4 (GnT IV) EC 2.4.1.145
17. human cDNA, Yoshida et al (1998) Glycoconjugate Journal 15:1115-1123
18. bovine cDNA, Minowa et al., European Patent EP 0 905 232
- 30 30  $\beta$ 1,6 (GnT V) EC 2.4.1.155

19. human cDNA, Saito et al (1994) *Biochem. Biophys. Res. Commun.* 198:318-327
20. rat cDNA, Shoreibah et al (1993) *J. Biol. Chem.* 268:15381-15385  
 $\beta$ 1.4 Galactosyltransferase, EC 2.4.1.90 (LacNAc synthetase) EC
- 5 2.4.1.22 (lactose synthetase)
21. bovine cDNA, D'Agostaro et al (1989) *Eur. J. Biochem.* 183:211-217
22. bovine cDNA (partial), Narimatsu et al (1986) *Proc. Natl. Acad. Sci. USA* 83:4720-4724
23. bovine cDNA (partial), Masibay & Qasba (1989) *Proc. Natl. Acad. Sci. USA* 86:5733-5377
- 10 24. bovine cDNA (5' end), Russo et al (1990) *J. Biol. Chem.* 265:3324
25. chicken cDNA (partial), Ghosh et al (1992) *Biochem. Biophys. Res. Commun.* 1215-1222
26. human cDNA, Masri et al (1988) *Biochem. Biophys. Res. Commun.* 157:657-663
- 15 27. human cDNA, (HeLa cells) Watzele & Berger (1990) *Nucl. Acids Res.* 18:7174
28. human cDNA, (partial) Uejima et al (1992) *Cancer Res.* 52:6158-6163
- 20 29. human cDNA, (carcinoma) Appert et al (1986) *Biochem. Biophys. Res. Commun.* 139:163-168
30. human gene, Mengle-Gaw et al (1991) *Biochem. Biophys. Res. Commun.* 176:1269-1276
31. murine cDNA, Nakazawa et al (1988) *J. Biochem.* 104:165-168
- 25 32. murine cDNA, Shaper et al (1988) *J. Biol. Chem.* 263:10420-10428
33. murine cDNA (novel), Uehara & Muramatsu unpublished
34. murine gene, Hollis et al (1989) *Biochem. Biophys. Res. Commun.* 162:1069-1075
35. rat protein (partial), Bendiak et al (1993) *Eur. J. Biochem.* 216:405-417  
2.3-Sialyltransferase, (ST3Gal II) (N-linked) (Gal-1,3/4-GlcNAc) EC  
2.4.99.6

36. human cDNA, Kitagawa & Paulson (1993) Biochem. Biophys. Res. Commun. 194:375-382

37. rat cDNA, Wen et al (1992) J. Biol. Chem. 267:21011-21019  
2,6-Sialyltransferase, (ST6Gal I) EC 2.4.99.1

5 38. chicken, Kurosawa et al (1994) Eur. J. Biochem 219:375-381

39. human cDNA (partial), Lance et al (1989) Biochem. Biophys. Res. Commun. 164:225- 232

40. human cDNA, Grundmann et al (1990) Nucl. Acids Res. 18:667

41. human cDNA, Zettlmeisl et al (1992) Patent EPO475354-A/3

10 42. human cDNA, Stamenkovic et al (1990) J. Exp. Med. 172:641-643  
(CD75)

43. human cDNA, Bast et al (1992) J. Cell Biol. 116:423-435

44. human gene (partial), Wang et al (1993) J. Biol. Chem. 268:4355-4361

15 45. human gene (5' flank), Aasheim et al (1993) Eur. J. Biochem. 213:467-475

46. human gene (promoter), Aas-Eng et al (1995) Biochim. Biophys. Acta 1261:166-169

47. mouse cDNA, Hamamoto et al (1993) Bioorg. Med. Chem. 1:141-145

20 48. rat cDNA, Weinstein et al (1987) J. Biol. Chem. 262:17735-17743

49. rat cDNA (transcript fragments), Wang et al (1991) Glycobiology 1:25-31, Wang et al (1990) J. Biol. Chem. 265:17849-17853

50. rat cDNA (5' end), O'Hanlon et al (1989) J. Biol. Chem. 264:17389-17394; Wang et al (1991) Glycobiology 1:25-31

25 51. rat gene (promoter), Svensson et al (1990) J. Biol. Chem. 265:20863-20688

52. rat mRNA (fragments), Wen et al (1992) J. Biol. Chem. 267:2512-2518

30 Additional methods and reagents which can be used in the methods for modifying the glycosylation are described in the literature, such as U.S.

Patent No. 5,955,422, U.S. Patent No. 4,775,622, U.S. Patent No. 6,017,743,  
U.S. Patent No. 4,925,796, U.S. Patent No. 5,766,910, U.S. Patent No.  
5,834,251, U.S. Patent No. 5,910,570, U.S. Patent No. 5,849,904, U.S.  
Patent No. 5,955,347, U.S. Patent No. 5,962,294, U.S. Patent No. 5,135,854,  
5 U.S. Patent No. 4,935,349, U.S. Patent No. 5,707,828, and U.S. Patent No.  
5,047,335.

Appropriate yeast expression systems can be obtained from sources such as the American Type Culture Collection, Rockville, MD. Vectors are commercially available from a variety of sources.

I claim:

1. A method for producing glycoproteins having carbohydrate structures similar to those produced by human cells in a lower eukaryote comprising providing a unicellular or multicellular fungal host, which does not express one or more enzymes involved in production of high mannose structures, and

introducing into the host one or more enzymes for production of a carbohydrate structure selected from the group consisting of  $\text{Man}_5\text{GlcNAc}_2$ ,  $\text{Man}_8\text{GlcNAc}_2$  and  $\text{Man}_9\text{GlcNAc}_2$ , wherein the enzymes are selected to have optimal activity at the pH of the location in the host where the carbohydrate structure is produced or which are targeted to a subcellular location in the host where enzyme will have optimal activity to produce the carbohydrate structure.

2. The method of claim 1 wherein the host is deficient in the activity of one or more enzymes selected from the group consisting of mannosyltransferases and phosphomannosyltransferases.

3. The method of claim 2 wherein the host does not express an enzyme selected from the group consisting of 1,6 mannosyltransferase, 1,3 mannosyltransferase, and 1,2 mannosyltransferase.

4. The method of claim 1 wherein the host is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Candida albicans*, *Aspergillus nidulans*, and *Trichoderma reesei*.

5. The method of claim 2 wherein the host is an OCH1 mutant of *P. pastoris*.

6. The method of claim 1 comprising introducing into the host a nucleotide molecule encoding one or more mannosidases involved in the production of  $\text{Man}_5\text{GlcNAc}_2$  from  $\text{Man}_8\text{GlcNAc}_2$  or  $\text{Man}_9\text{GlcNAc}_2$ .

7. The method of claim 6 where the at least one mannosidase has a pH optimum within 1.4 pH units of the average pH optimum of other

representative enzymes in the organelle in which the mannosidase is localized, or having optimal activity at a pH between 5.1 and 8.0.

8. The method of claim 7 wherein the mannosidase enzyme has optimal activity at a pH between 5.9 and 7.5.

9. The method of claim 8 wherein the mannosidase enzyme is an  $\alpha$ -1,2-mannosidase derived from mouse, human, *Lepidoptera*, *Aspergillus nidulans*, or *Bacillus* sp.

10. The method of claim 1 comprising providing a host that is able to form  $\text{Man}_5\text{GlcNAc}_2$  structures, displaying GnT I activity and having UDP-Gn transporter activity.

11. The method of claim 1 comprising providing a host which has a UDP specific diphosphatase activity.

12. The method of claim 1 comprising introducing into the host one or more enzymes selected from the group consisting of mannosidases, glycosyltransferases and glycosidases, wherein the enzymes are targeted to the endoplasmic reticulum, the early, medial, late Golgi or the trans Golgi network.

13. The method of claim 12 wherein the mannosidase enzyme is predominantly localized in the Golgi apparatus or the endoplasmic reticulum.

14. The method of claim 12 wherein the enzymes are localized by forming a fusion protein between a catalytic domain of the enzyme and a chimeric localization region encoded by at least one genetic construct formed by the in-frame ligation of a DNA fragment encoding a cellular targeting signal peptide with a DNA fragment encoding a glycosylation enzyme or catalytically active fragment thereof

15. The method of claim 14 comprising providing a chimeric localization region from an enzyme selected from the group of mannosyltransferases, diphosphatases, proteases, GnT I, GnT II, GnT III, GnT IV, GnT V, GnT VI, GalT, FT, and ST.

16. The method of claim 14 providing a catalytic domain encoding a glycosidase or glycosyltransferase selected from the group consisting of GnT I, GnT II, GnT III, GnT IV, GnT V, GnT VI, GalT, Fucosyltransferase and ST, that has a pH optimum within 1.4 pH units of the average pH optimum of other

representative enzymes in the organelle in which the enzyme is localized, or having optimal activity at a pH between 5.1 and 8.0.

17. The method of claim 1 comprising introducing into the host nucleotide molecules encoding one or more enzymes selected from the group of nucleoside sugar transporters consisting of UDP-GlcNAc transferase, UDP-galactosyltransferase, GDP-fucosyltransferase, CMP-sialyltransferase, UDP-GlcNAc transporter, UDP-galactose transporter, GDP-fucose transporter, CMP-sialic acid transporter, and nucleotide diphosphatases.

18. The method of claim 17 comprising genetically engineering the fungal strain to remove UDP or GDP by action of a diphosphatase.

19. The method of claim 1 wherein the glycoprotein includes *N*-glycans of which greater than 27 mole percent comprise fewer than six mannose residues.

20. The method of claim 1 wherein the glycoprotein comprises one or more sugars selected from the group consisting of galactose, sialic acid, and fucose.

21. The method of claim 1 wherein the glycoprotein comprises at least one oligosaccharide branch comprising the structure NeuNAc-Gal-GlcNAc-Man.

22. The method of claim 1 wherein the glycoprotein comprises *N*-glycans having fewer than four mannose residues.

23. The method of claim 1 wherein subsequent to isolation from the host, the glycoprotein is subjected to at least one further glycosylation or carboxylation reaction *in vitro*.

24. The method of claim 1 comprising the steps of

(a) providing a DNA library comprising at least two genes encoding exogenous glycosylation enzymes;

(b) transforming the host with the library to produce a genetically mixed population expressing at least two distinct exogenous glycosylation enzymes; and

(c) selecting from the population a host producing the desired glycosylation phenotype.

25. The method of claim 24 wherein the host is transformed two or more times with the library prior to the selection of a desired glycosylation phenotype.

26. The method of claim 24 wherein the library comprises at least one wild-type gene encoding a glycosylation enzyme.

27. The method of claim 24 wherein the library comprises at least one synthetic gene encoding a glycosylation enzyme.

28. The method of claim 24 wherein the library comprises at least one gene previously subjected to a technique selected from the list: gene shuffling, *in vitro* mutagenesis, and error-prone polymerase chain reaction.

29. The method of claim 24 wherein the library comprises at least one genetic construct formed by the in-frame ligation of a DNA fragment encoding a cellular targeting signal peptide with a DNA fragment encoding a glycosylation enzyme or catalytically active fragment thereof.

30. The method of claim 29 wherein a DNA fragment encodes an activity selected from the group consisting of mannosidase, UDP-GlcNAc transferase, UDP-galactosyltransferase, and CMP-sialyltransferase and the cellular targeting signal peptide is predominantly localizes the enzyme in an organelle selected from the group consisting of endoplasmic reticulum, cis Golgi, medial Golgi, and trans Golgi.

31. The method of claim 24 wherein the selection comprises the step of analyzing a glycosylated protein or isolated *N*-glycan by one or more methods selected from the group consisting of mass spectroscopy, liquid chromatography, characterizing cells using a fluorescence activated cell sorter, spectrophotometer, fluorimeter, or scintillation counter, exposing host cells to a lectin or antibody having a specific affinity for a desired oligosaccharide moiety, and exposing cells to a cytotoxic or radioactive molecule selected from the group consisting of sugars, antibodies, and lectins.

32. The host produced by the method of any of claims 1-31.

33. The glycoprotein produced by the method of any of claims 1-31.

34. The library of claims 24-30.

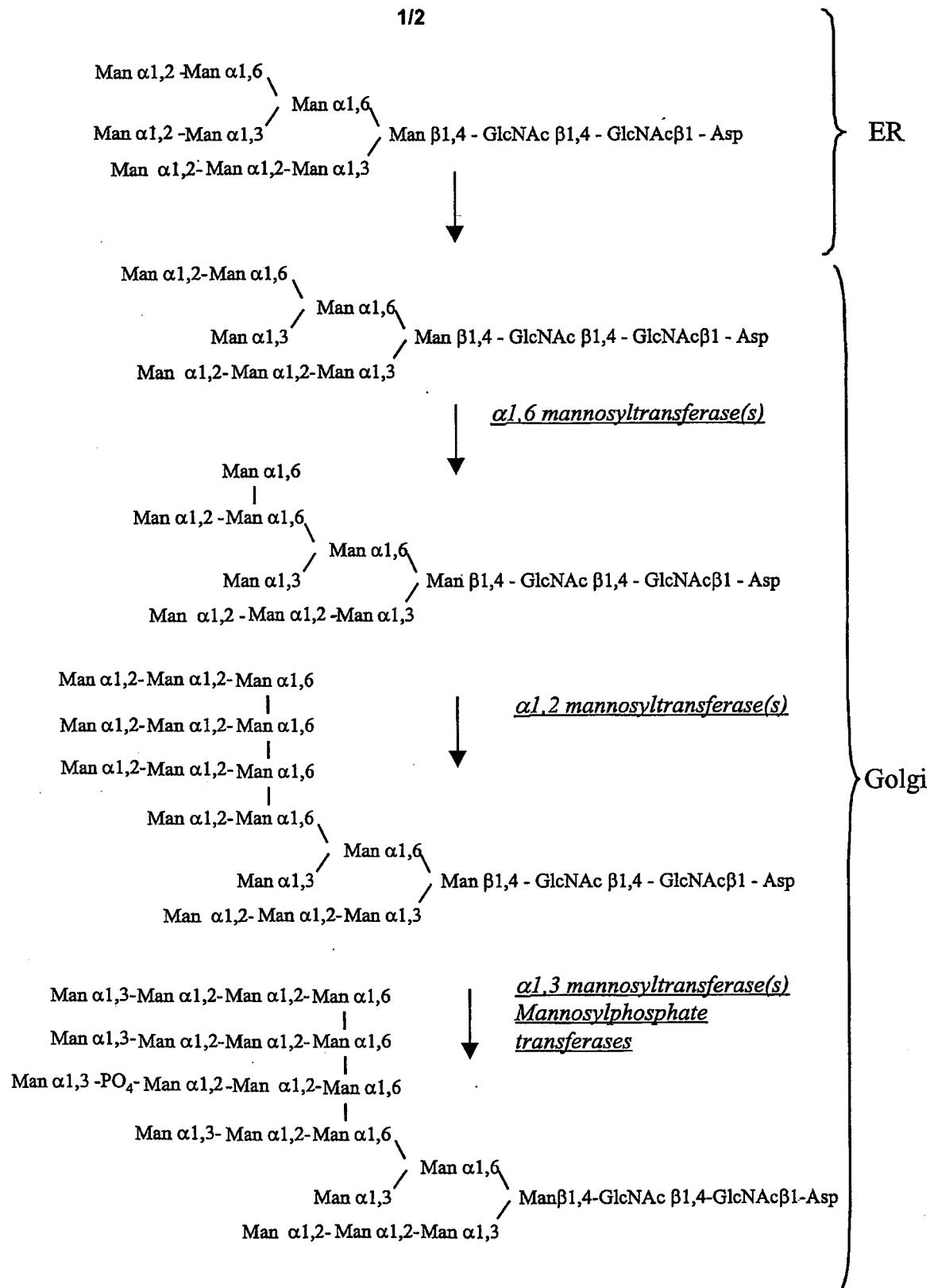


Figure 1.A

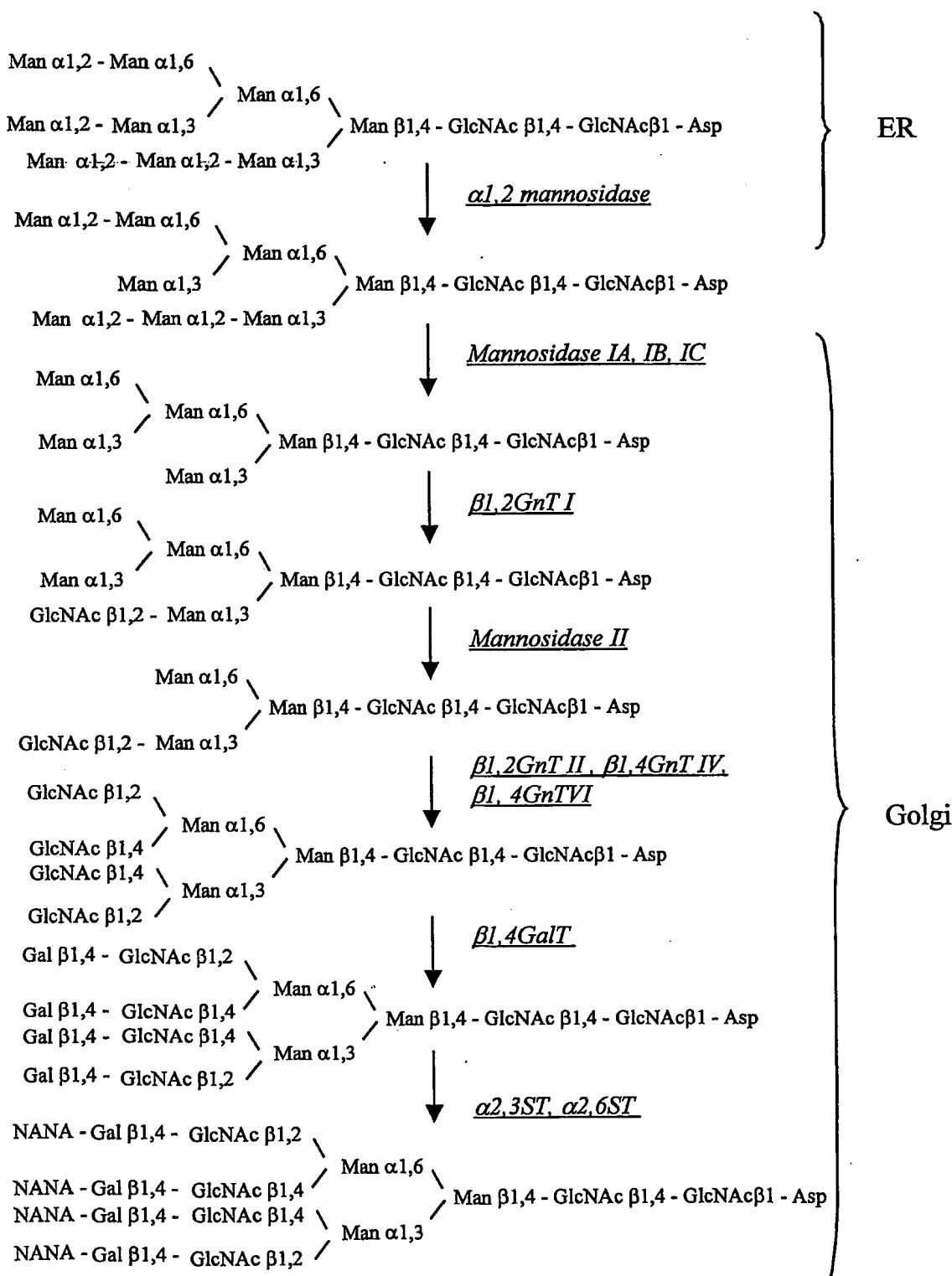


Figure 1.B